16. Path-following methods

- central path
- short-step barrier method
- predictor-corrector method
Introduction

primal-dual pair of conic LPs

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax \preceq b
\end{align*}
\]
\[
\begin{align*}
\text{maximize} & \quad -b^T z \\
\text{subject to} & \quad A^T z + c = 0 \\
& \quad z \succeq^* 0
\end{align*}
\]

- \(A \in \mathbb{R}^{m \times n}\) with \(\text{rank}(A) = n\)
- inequalities are with respect to proper cone \(K\) and its dual cone \(K^*\)
- we will assume primal and dual problem are strictly feasible

this lecture

- feasible methods that follow the central path to find the solution
- complexity analysis based on theory of self-concordant functions
Outline

• central path
• short-step barrier method
• predictor-corrector method
Barrier for the feasible set

**definition:** as a barrier function for the feasible set we will use

$$\psi(x) = \phi(b - Ax)$$

where $\phi$ is a $\theta$-normal barrier for $K$

**notation** (in this lecture): $\|v\|_{x^*} = (v^T \nabla^2 \psi(x)^{-1} v)^{1/2}$

**properties**

- $\psi$ is self-concordant with domain $\{x \mid Ax < b\}$
- Newton decrement of $\psi$ is bounded by $\sqrt{\theta}$, i.e.,

$$\|\nabla \psi(x)\|_{x^*}^2 = \nabla \psi(x)^T \nabla^2 \psi(x)^{-1} \nabla \psi(x) \leq \theta \quad \forall x \in \text{dom } \psi$$

(proof on next page)
proof of bound on Newton decrement

- gradient and Hessian of $\psi$ are (with $s = b - Ax$)

$$\nabla \psi(x) = -A^T \nabla \phi(s), \quad \nabla^2 \psi(x) = A^T \nabla^2 \phi(s) A$$

- from page 15-24, $\nabla \phi(s)^T \nabla^2 \phi(s)^{-1} \nabla \phi(s) = \theta$; therefore

$$\begin{align*}
\nabla \psi(x)^T \nabla^2 \psi(x)^{-1} \nabla \psi(x) &= \sup_v (-v^T \nabla^2 \psi(x) v + 2 \nabla \psi(x)^T v) \\
&= \sup_v (-v^T \nabla^2 \phi(s) (Av) - 2 \nabla \phi(s)^T Av) \\
&\leq \sup_w (-w^T \nabla^2 \phi(s) w + 2 \nabla \phi(s)^T w) \\
&= \nabla \phi(s)^T \nabla^2 \phi(s)^{-1} \nabla \phi(s) \\
&= \theta
\end{align*}$$
Central path

**definition:** the set of minimizers \( x^*(t) \), for \( t > 0 \), of

\[
tc^T x + \psi(x) = tc^T x + \phi(b - Ax)
\]

**optimality conditions**

\[
A^T \nabla \phi(s) = tc, \quad s = b - Ax
\]

- implies that \( z = -(1/t) \nabla \phi(s) \) is strictly dual feasible
- by weak duality,

\[
c^T x^*(t) - p^* \leq c^T x + b^T z = z^T s = \frac{\theta}{t}
\]

hence, \( c^T x^*(t) \rightarrow p^* \) as \( t \rightarrow \infty \)
Existence and uniqueness

centering problem

\[
\begin{align*}
\text{minimize} & \quad tc^T x + \phi(s) \\
\text{subject to} & \quad Ax + s = b
\end{align*}
\]

Lagrange dual (with dual cone barrier $\phi_*$ of page 15-27)

\[
\begin{align*}
\text{maximize} & \quad -tb^T z - \phi_*(z) + \theta \log t \\
\text{subject to} & \quad A^T z + c = 0
\end{align*}
\]

- strictly feasible $z$ for dual conic LP is feasible for dual centering problem
- if dual conic LP is strictly feasible, $tc^T x + \phi(b - Ax)$ is bounded below
- from self-concordance theory (p.15-12), $x^*(t)$ exists and is unique
Dual points in neighborhood of central path

**Newton step** $\Delta x$ for $tc^T x + \psi(x) = tc^T x + \phi(b - Ax)$

- satisfies Newton equation
  \[
  A^T \nabla^2 \phi(s) A \Delta x = -tc + A^T \nabla \phi(s), \quad s = b - Ax
  \]

- Newton decrement is $\lambda_t(x) = (\Delta x^T \nabla^2 \psi(x) \Delta x)^{1/2}$

**dual feasible point:** define

\[
  z = -\frac{1}{t} \left( \nabla \phi(s) - \nabla^2 \phi(s) A \Delta x \right)
\]

- satisfies $A^T z + c = 0$ by definition
- satisfies $z \succ_0 0$ if $\lambda_t(x) < 1$ (see next page)
\textit{proof.} \( z \succ_\ast 0 \) follows from Dikin ellipsoid theorem

- Newton decrement is

\[
\lambda_t(x)^2 = \Delta x^T \nabla^2 \psi(x) \Delta x \\
= \Delta x^T A^T \nabla^2 \phi(s) A \Delta x \\
= v^T \nabla^2 \phi(s)^{-1} v
\]

where \( v = \nabla^2 \phi(s) A \Delta x \)

- define \( u = -\nabla \phi(s) \); then \( \nabla^2 \phi_*(u) = \nabla^2 \phi(s)^{-1} \) (see p.15-28) and

\[
\lambda_t(x)^2 = v^T \nabla^2 \phi_*(u) v
\]

- by Dikin ellipsoid theorem \( \lambda_t(x) < 1 \) implies

\[
u + v = -\nabla \phi(s) + \nabla^2 \phi(s) A \Delta x \succ_\ast 0
\]
Duality gap in neighborhood of central path

\[ c^T x - p^* \leq \left( 1 + \frac{\lambda_t(x)}{\sqrt{\theta}} \right) \frac{\theta}{t} \quad \text{if } \lambda_t(x) < 1 \]

• from weak duality, using the dual point \( z \) on page 16-7

\[ s^T z = \frac{1}{t} \left( \theta - s^T \nabla^2 \phi(s) A \Delta x \right) \]

\[ \leq \frac{1}{t} \left( \theta + \| \nabla^2 \phi(s)^{1/2} s \|_2 \| \nabla^2 \phi(s)^{1/2} A \Delta x \|_2 \right) \]

\[ = \frac{\theta + \sqrt{\theta} \lambda_t(x)}{t} \]

• implies \( c^T x - p^* \leq 2\theta/t \), since \( \theta \geq 1 \) holds for any \( \theta \)-normal barrier \( \phi \)

\( (\phi \text{ is unbounded below, so its Newton decrement } \sqrt{\theta} \geq 1 \text{ everywhere}) \)
Outline

• central path

• short-step barrier method

• predictor-corrector method
Short-step methods

general idea: keep the iterates in the region of quadratic convergence for
\[ tc^T x + \psi(x), \]

by limiting the rate at which \( t \) is increased (hence, ‘short-step’)

quadratic convergence results (from self-concordance theory)

• if \( \lambda_t(x) \leq 1/4 \), a full Newton step gives \( \lambda_t(x^+) \leq 2\lambda_t(x)^2 \)
• started at a point with \( \lambda_t(x) \leq 1/4 \), an accuracy \( \varepsilon_{\text{cent}} \) is reached in
\[
\log_2 \log_2 (1/\varepsilon_{\text{cent}}) \text{ iterations}
\]

for practical purposes this is a constant (4–6 for \( \varepsilon_{\text{cent}} \approx 10^{-5} \ldots 10^{-20} \))
Short-step method with exact centering

simplifying assumptions:

• \( x^*(t) \) is computed exactly

• a central point \( x^*(t_0) \) is given

algorithm: define a tolerance \( \epsilon \in (0, 1) \) and parameter

\[
\mu = 1 + \frac{1}{4\sqrt{\theta}}
\]

starting at \( t = t_0 \), repeat until \( \theta/t \leq \epsilon \):

• compute \( x^*(\mu t) \) by Newton’s method started at \( x^*(t) \)

• set \( t := \mu t \)
Newton iterations for recentering

Newton decrement at $x = x^*(t)$ for new value $t^+ = \mu t$ is

$$
\lambda_{t^+}(x) = \|\mu tc + \nabla \psi(x)\|_{x^*} \\
= \|\mu(tc + \nabla \psi(x)) - (\mu - 1)\nabla \psi(x)\|_{x^*} \\
= (\mu - 1)\|\nabla \psi(x)\|_{x^*} \\
\leq (\mu - 1)\sqrt{\theta} \\
= 1/4
$$

- line 3 follows because $tc + \nabla \psi(x) = 0$ for $x = x^*(t)$
- line 4 follows from $\|\nabla \psi(x)\|_{x^*} \leq \sqrt{\theta}$ (see page 16-3)

**conclusion**

#iterations to compute $x^*(t^+) \text{ from } x^*(t)$ is bounded by a small constant
Iteration complexity

number of outer iterations: $t^{(k)} = \mu^k t_0 \geq \theta/\epsilon$ when

$$k \geq \frac{\log(\theta/(\epsilon t_0))}{\log \mu}$$

cumulative number of Newton iterations

$$O\left(\sqrt{\theta} \log \left(\frac{\theta}{\epsilon t_0}\right)\right)$$

(we used $\log \mu \geq (\log 2)/(4\sqrt{\theta})$ by concavity of $\log(1 + u)$)

- multiply by flops per iteration to get polynomial worst-case complexity
- $\sqrt{\theta}$ dependence is lowest known complexity for interior-point methods
Short-step method with inexact centering

**Improvements** of short-step method with exact centering

- keep iterates in region of quadratic region, but avoid complete centering
- at each iteration: make small increase in $t$, followed by *one* Newton step

**Algorithm:** define a tolerance $\epsilon \in (0, 1)$ and parameters

$$\beta = \frac{1}{8}, \quad \mu = 1 + \frac{1}{1 + 8\sqrt{\theta}}$$

- select $x$ and $t$ with $\lambda_t(x) \leq \beta$
- repeat until $2\theta/t \leq \epsilon$:

$$t := \mu t, \quad x := x - \nabla^2 \psi(x)^{-1} (tc + \nabla \psi(x))$$
Newton decrement after update

we first show that $\lambda_t(x) \leq \beta$ at the end of each iteration

• if $\lambda_t(x) \leq \beta$ and $t^+ = \mu t$, then

\[
\lambda_{t^+}(x) = \|t^+ c + \nabla \psi(x)\|_{x^*}
\]
\[
= \|\mu(tc + \nabla \psi(x)) - (\mu - 1)\nabla \psi(x)\|_{x^*}
\]
\[
\leq \mu \|tc + \nabla \psi(x)\|_{x^*} + (\mu - 1)\|\nabla \psi(x)\|_{x^*}
\]
\[
\leq \mu \beta + (\mu - 1)\sqrt{\theta}
\]
\[
= \frac{1}{4}
\]

• from theory of Newton’s method for s.c. functions (p.15-16)

\[
\lambda_{t^+}(x^+) \leq 2\lambda_{t^+}(x)^2 \leq \frac{1}{8} = \beta
\]
Iteration complexity

• from page 16-9, stopping criterion implies $c^T x - p^* \leq \epsilon$

• stopping criterion is satisfied when

$$\frac{t^{(k)}}{t_0} = \mu^k \geq \frac{2\theta}{\epsilon t_0}, \quad k \geq \frac{\log(2\theta/(\epsilon t_0))}{\log \mu}$$

• taking the logarithm on both sides gives an upper bound of

$$O \left( \sqrt{\theta} \log \left( \frac{\theta}{\epsilon t_0} \right) \right) \text{ iterations}$$

(using $\log \mu \geq \log 2/(1 + 8\sqrt{\theta})$)
• central path

• short-step barrier method

• predictor-corrector method
Predictor-corrector methods

**short-step methods**

- stay in narrow neighborhood of central path (defined by limit on $\lambda_t$)
- make small, fixed increases $t^+ = \mu t$

as a result, quite slow in practice

**predictor-corrector method**

- select new $t$ using a linear approximation to central path (‘predictor’)
- recenter with new $t$ (‘corrector’)

allows faster and ‘adaptive’ increases in $t$
Global convergence bound for centering problem

\[ \text{minimize } f_t(x) = tc^T x + \phi(b - Ax) \]

**Convergence result** (damped Newton algorithm of p.15-11 started at \( x \))

\[ \# \text{iterations} \leq \frac{f_t(x) - \inf_u f_t(u)}{\omega(\eta)} + \log_2 \log_2(1/\epsilon_{\text{cent}}) \]

- \( \epsilon_{\text{cent}} \) is accuracy in centering; \( \eta \in (0, 1/4] \); \( \omega(\eta) = \eta - \log(1 + \eta) \)
- for practical purposes, second term is a small constant
- first term depends on unknown optimal value \( \inf_u f_t(u) \)
Bound from duality

dual centering problem (see p.16-6)

\[
\begin{align*}
\text{maximize} & \quad -tb^T z - \phi_*(z) + \theta \log t \\
\text{subject to} & \quad A^T z + c = 0
\end{align*}
\]

strictly feasible $z$ provides lower bound on $\inf_u f_t(u)$:

\[
\inf_u f_t(u) \geq -tb^T z - \phi_*(z) + \theta \log t
\]

bound on centering cost: $f_t(x) - \inf_u f_t(u) \leq V_t(x, s, z)$ where

\[
V_t(x, s, z) = t(c^T x + b^T z) + \phi(s) + \phi_*(z) - \theta \log t
\]

\[
= ts^T z + \phi(s) + \phi_*(z) - \theta \log t
\]
Potential function

**Definition** (for strictly feasible $x, s, z$)

$$
\Psi(x, s, z) = \inf_t V_t(x, s, z)
$$

$$
= \theta \log \frac{s^T z}{\theta} + \phi(s) + \phi^*(z) + \theta
$$

(optimal $t$ is $t = \arg\min_t V_t(x, s, z) = \theta / s^T z$)

**Properties**

- homogeneous of degree zero: $\Psi(\alpha x, \alpha s, \alpha z) = \Psi(x, s, z)$ for $\alpha > 0$
- nonnegative for all strictly feasible $x, s, z$
- zero only if $x, s, z$ are centered

can be used as a *global* proximity measure
Tangent to central path

central path equation

\[
\begin{bmatrix}
0 \\
s^*(t)
\end{bmatrix} = \begin{bmatrix}
0 & A^T \\
-A & 0
\end{bmatrix} \begin{bmatrix}
x^*(t) \\
z^*(t)
\end{bmatrix} + \begin{bmatrix}
c \\
b
\end{bmatrix}
\]

\[
z^*(t) = -\frac{1}{t} \nabla \phi(s^*(t))
\]

derivatives \( \dot{x} = dx^*(t)/dt, \ \dot{s} = ds^*/dt, \ \dot{z} = dz^*(t)/dt \) satisfy

\[
\begin{bmatrix}
0 \\
\dot{s}
\end{bmatrix} = \begin{bmatrix}
0 & A^T \\
-A & 0
\end{bmatrix} \begin{bmatrix}
\dot{x} \\
\dot{z}
\end{bmatrix}
\]

\[
\dot{z} = -\frac{1}{t} z^*(t) - \frac{1}{t} \nabla^2 \phi(s^*(t)) \dot{s}
\]

tangent direction: defined as \( \Delta x_t = t \dot{x}, \ \Delta s_t = t \dot{s}, \ \Delta z_t = t \dot{z} \)
Predictor equations

with \( x = x^*(t) \), \( s = s^*(t) \), \( z = z^*(t) \)

\[
\begin{bmatrix}
(1/t)\nabla^2\phi(s) & 0 & I \\
0 & 0 & A^T \\
-I & -A & 0
\end{bmatrix}
\begin{bmatrix}
\Delta s_t \\
\Delta x_t \\
\Delta z_t
\end{bmatrix}
= \begin{bmatrix}
-z \\
0 \\
0
\end{bmatrix}
\]

(1)

equivalent equations

\[
\begin{bmatrix}
I & 0 & (1/t)\nabla^2\phi_*(z) \\
0 & 0 & A^T \\
-I & -A & 0
\end{bmatrix}
\begin{bmatrix}
\Delta s_t \\
\Delta x_t \\
\Delta z_t
\end{bmatrix}
= \begin{bmatrix}
-s \\
0 \\
0
\end{bmatrix}
\]

(2)

equivalence follows from primal-dual relations on central path

\[
z = -\frac{1}{t}\nabla\phi(s), \quad s = -\frac{1}{t}\nabla\phi_*(z), \quad \frac{1}{t}\nabla^2\phi(s) = t\nabla^2\phi_*(z)^{-1}
\]
Properties of tangent direction

- from 2nd and 3rd block in (1): $\Delta s^T_t \Delta z_t = 0$
- from first block in (1) and $\nabla^2 \phi(s)s = -\nabla \phi(s)$:

$$s^T \Delta z_t + z^T \Delta s_t = -s^T z$$

- hence, gap in tangent direction is

$$(s + \alpha \Delta s_t)^T (z + \alpha \Delta z_t) = (1 - \alpha) s^T z$$

- from first block in (1)

$$\|\Delta s_t\|_s^2 = \Delta s_t^T \nabla^2 \phi(s) \Delta s_t = -tz^T \Delta s_t$$

- similarly, from first block in (2)

$$\|\Delta z_t\|_z^2 = \Delta z_t^T \nabla^2 \phi_*(z) \Delta z_t = -ts^T \Delta z_t$$
Predictor-corrector method with exact centering

simplifying assumptions: exact centering, a central point $x^*(t_0)$ is given

algorithm: define tolerance $\epsilon \in (0, 1)$, parameter $\beta > 0$, and set

$$t := t_0, \quad (x, s, z) := (x^*(t_0), s^*(t_0), z^*(t_0))$$

repeat until $\theta/t \leq \epsilon$:

- compute tangent direction $(\Delta x_t, \Delta s_t, \Delta z_t)$ at $(x, s, z)$
- set $(x, s, z) := (x, s, z) + \alpha(\Delta x_t, \Delta s_t, \Delta z_t)$ with $\alpha$ determined from

$$\Psi(x + \alpha \Delta x_t, s + \alpha \Delta s_t, z + \alpha \Delta z_t) = \beta$$

- set $t := \theta/(s^T z)$ and compute $(x, s, z) := (x^*(t), s^*(t), z^*(t))$
Iteration complexity

potential function in tangent direction (proof on next page)

\[ \Psi(x + \alpha \Delta x_t, s + \alpha \Delta s_t, z + \alpha \Delta s_t) \leq \omega^*(\alpha \sqrt{\theta}) \]

\[ = -\alpha \sqrt{\theta} - \log(1 - \alpha \sqrt{\theta}) \]

lower bound on predictor step length: since \( \omega^* \) is an increasing function

\[ \alpha \geq \gamma / \sqrt{\theta} \quad \text{where} \quad \omega^*(\gamma) = \beta \]

reduction in duality gap after one predictor/corrector cycle

\[ t/t^+ = 1 - \alpha \leq 1 - \gamma / \sqrt{\theta} \leq \exp(-\gamma / \sqrt{\theta}) \]

cumulative Newton iterations: \( t^{(k)} \geq \theta / \epsilon \) after

\[ O \left( \sqrt{\theta} \log \left( \theta / (t_0 \epsilon) \right) \right) \quad \text{Newton iterations} \]
proof of upper bound on $\Psi$ (with $s^+ = s + \alpha \Delta s_t$, $z^+ = z + \alpha \Delta z_t$)

- bounds on $\phi(s^+)$ and $\phi_*(z^+)$: from the inequality on page 15-8,

$$\phi(s^+) - \phi(s) \leq \alpha \nabla \phi(s)^T \Delta s_t + \omega^*(\alpha \|\Delta s_t\|_s)$$
$$= -\alpha tz^T \Delta s_t + \omega^*(\alpha \|\Delta s_t\|_s)$$

$$\phi_*(z^+) - \phi_*(z) \leq \alpha \nabla \phi(z)^T \Delta z_t + \omega^*(\alpha \|\Delta z_t\|_z)$$
$$= -\alpha ts^T \Delta z_t + \omega^*(\alpha \|\Delta z_t\|_z)$$

- add the inequalities and use properties on page 16-23

$$\phi(s^+) - \phi(s) + \phi_*(z^+) - \phi_*(z) \leq \alpha \theta + \omega^*(\alpha \|\Delta s_t\|_s) + \omega^*(\alpha \|\Delta z_t\|_z)$$
$$\leq \alpha \theta + \omega^*(\alpha (\|\Delta s_t\|_s^2 + \|\Delta z_t\|_z^2)^{1/2})$$
$$= \alpha \theta + \omega^*(\alpha \sqrt{\theta})$$

- since $(s^+)^T z^+ = (1 - \alpha) s^T z$,

$$\Psi(x^+, s^+, z^+) \leq \theta \log(1 - \alpha) + \alpha \theta + \omega^*(\alpha \sqrt{\theta}) \leq \omega^*(\alpha \sqrt{\theta})$$
References